DOCUMENT RESUME

ED 051 279 TM 000 593

AUTHOR	Huberty, Carl J.
TITLE	On the Variable Selection Problem in Multiple Group Discriminant Analysis.
PUB DATE	Feb 71
NOTE	39p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Cluster Analysis, Correlation, Criteria, *Discriminant Analysis, Factor Analysis, *Mathematical Models, Mathematics, *Multiple Regression Analysis, *Predictor Variables, Research Methodology, *Statistical Analysis, Tests of Significance

ABSTRACT

This study was concerned with various schemes for reducing the number of variables in a multivariate analysis. Two sets of illustrative data were used; the numbers of criterion groups were 3 and 5. The proportion of correct classifications was employed as an index of discriminatory power of each subset of variables selected. Of the four procedures using indices that order the variables with respect to contribution to discrimination, the (forward) stepwise procedure yielded the best results. Of the two schemes involving dimensional analysis, that which uses correlations of scores on variables with high maximum likelihood factor loadings against discriminant scores appeared more attractive. (Author)

ON THE VARIABLE SELECTION PROBLEM IN MULTIPLE GROUP DISCRIMINANT ANALYSIS

CARL J HUBERTY

UNIVERSITY OF GEORGIA

Presented at the Annual Meeting of the

American Educational Research Association

New York, February, 1971

Abstract

This study was concerned with various schemes for reducing the number of variables in a multivariate analysis. 2 sets of illustrative data were used the numbers of criterion groups were 3 and 5. The proportion of correct classifications was employed as an index of discriminatory power of each subset of variables selected. Of the 4 procedures using indices that order the variables with respect to contribution to discrimination, the (forward) stepwise procedure yielded the best results. Of the 2 schemes involving dimensional analysis, that which uses correlations of scores on variables with high maximum likelihood factor loadings against discriminant scores appeared more attractive.

ON THE VARIABLE SELECTION PROBLEM IN MULTIPLE GROUP DISCRIMINANT ANALYSIS

Carl J Huberty[1]

University of Georgia

In many situations involving multiple group discriminant analysis the investigator is presented with more variables than he would like and there arises the question of whether they are all necessary and, if not, which of them can be discarded. Having obtained weights which determine the discriminant scores, the investigator may ask if the data might not have been adequately explained by using only a subset of the original input variables. When an experimenter is confronted with such a problem he wants to include as many variables as possible so that reliable values may be determined, and yet as few as possible so as to keep the costs involved in obtaining information at a minimum.

For the case of a discriminatory analysis involving more than two criterion populations, no known optimizing procedure has yet been developed to reduce the number of discriminator variables. (Optimum in the sense that the variables selected would lead to a maximum amount of separation among the groups for that number of variables.) There is a dearth of literature concerning the reduction in the number of variables in discriminatory analysis. To date few studies have used any selection procedures in actual research and fewer still have subjected any methods to empirical comparisons. Until recently, of course--i.e., before electronic computers became readily available--the computations involved in discriminatory analysis and the deletion of variables proved to be a task of such herculean proportions that an investigator could hardly be blamed for not planning his study to include such an analysis.

Quenoville (1949a) developed a computational scheme for deleting variables when the underlying model was that of Fisher's, in which, for more than two criterion groups, collinearity of population means is assumed, and thus only one discriminant function is obtained. His scheme involves a first approximation to the eigenvector associated with the largest corresponding eigenvalue, and a process of successive iterations which leads to convergence of the discriminant function coefficients. Thus, he arrived at a procedure whereby the initial analysis is used to avoid the complete recalculation of a new discriminant function when potential discriminators are discarded. In a second article (1949b) he eliminated the difficulty of taking special action to prevent the coefficients from converging to zero.

Rao (1952, p 253; 1965, p. 482) presents a statistic which may be used as a criterion for discarding p-q potential discriminators. The statistic

$$\frac{N_1 + N_2 - p - 1}{p-q} \cdot \frac{N_1 N_2 (D_p^2 - D_q^2)}{(N_1 + N_2)(N_1 + N_2 - 2) + N_1 N_2 D_q^2},$$

where D_p^2 is Mahalanobis' estimate of the distance between populations 1 and 2 based on the original set of p measures, has an approximate F distribution with (p-q) and $(N_1 + N_2 - p - 1)$ degrees of freedom. This statistic may be employed to test the hypothesis that (p-q) measures on the variables do not provide additional discrimination. This test is applicable for only two p-variable populations at at a time (from which random samples of size N_1 and N_2 are selected).

Collier (1963) shows that since there is a direct relationship between regression analysis and discriminant analysis for the two-group case, methods for deleting variables in regression analysis could be carried over to discriminant analysis. He displayed the equivalence between the regression test (involving the

multiple correlation coefficient) and the above statistic of Rao's for testing the contribution of additional variables.

Cochran and Bliss (1948) employ a covariance technique (for use with two criterion populations) to determine if a particular discriminator, or subset of discriminators, contribute anything to the adjusted discriminant. In this analysis the questionable discriminators are treated as the covariates. An estimated standard error (based on an error mean square of an ANOVA) of a coefficient is found which is used in forming a t ratio. Thus, if omission of some discriminators seems warranted, these t ratios are relevant in deciding which variable to eliminate first.

Again considering the two-group case and a single discriminant, Kendall (1957, p. 163) gives the standard error (for large samples) of the coefficients in the linear discriminant. Thus, using a simple t test it can be determined if the variable corresponding to a sample coefficient can be discarded without a serious loss of discriminating power. Grimsley and Summers (1965) used such a technique in a study to delete three of four variables.

Some other attempts have been made to use methods by which the number of potential discriminators is reduced before computing a discriminant, on the grounds that their inclusion is unlikely to produce a material increase in power to discriminate between two groups. Horst and Smith (1950) were able to discard 7 of 18 original variables on which physical measures were used to differentiate between men of Japanese and Caucasian stock. Wallace and Travers (1938) used only 5 of more than 20 available variables in distinguishing successful from unsuccessful salesmen. In both situations the criterion used to retain a variable in the analysis was the significance (at the .01 level) rendered by a t test of difference between the two population means of the variable in question.

Cochran (1964) does not favor this latter test since the t-value is greatly influenced by the sizes of the samples that are selected for setting up the discriminant. On the contrary he states that ". . .if it is the fact that in practice most correlations are positive and modest in size, the analysis suggests that reliance on the value of Σd_i^2 in deciding whether to throw away a group of poor discriminators is unlikely to produce a serious mistake [p. 186]." The value d_i is defined to be the squared normalized difference of the means of variable i in the two criterion populations. Twelve examples from the literature indicated that it will usually be safe to reject a group of variables if the value of Σd_i^2 is small. Because of the relatively small sample sizes used in the examples cited, the conclusion may be somewhat questionable.

Procedures such as these have limited practicality in that they do not aid in determining which discriminators to delete or which to retain in a k-group (k 2) situation. Dunteman (1966) reduced the number of SVIE scales from 29 to 11 in a study where these scales were used to discriminate female students among five occupational groups He employed two different approaches to delete variables. The first approach was to use in the final analysis only those scales which had produced univariable F-ratios that were significant beyond the 01 level The second criterion used was that of selecting the variables which had relatively high weights on one or more of the discriminants obtained. Using the number of correct classifications as a criterion of effectiveness of the variables selected, his results showed that both of the variable reduction procedures resulted in about the same amount of efficiency as when including all of the original 29 variables. DeMann (1963) employed univariate chi-square tests to delete 12 out of 20 potential discriminators prior to his multivariate predictive analysis

The classical method of finding clusters of variables which are formed under certain consistent principles of classification is factor analysis, another

multivariate procedure. The purpose of factor analysis is the study of dependence patterns in the variables. This is accomplished by seeking artifical variables (factors) which may explain the dependence among the observable variables It is desirable to keep the number of such artifical variables as small as possible For greater ease in interpretation of multivariate data, factor analysis is sometimes used to reduce the number of variables. Troegh, et al (1957) followed this scheme in an attempt to determine the nature of masculinity and femininity in the preschool years; four criterion groups and 60 variables were involved A principal axis analysis of a pooled covariance matrix was performed which resulted in the extraction of 20 factors (approximately 100% of the estimated communality was accounted for). Two criteria dictated further reduction to four factors; these were rotated and mean factor scores were then estimated for all four factors for each criterion group These four "variables" were then subjected to a discriminant analysis. This approach to the variable problem is inappropriate since to obtain the factor scores it is necessary to still use the scores on all of the original variables Nonetheless, as we shall see later, factor analysis methodology may be employed as an aid in selecting a subset of discriminators.

Selection Procedures

The present study was concerned with various schemes for reducing the number of variables in a multiple group discriminant analysis design The analysis referred to is that of determining the eigenvectors associated with the eigenvalues (λ-values) obtained from the solution of the equation

$$|B - \lambda_1 W| = 0,$$

where

E = (pxp) pooled within-groups deviation score sum of squares and cross products (SSCP) matrix,

H = (pxp) among-groups deviation score SSCP matrix, and

I = (pxp) identity matrix.

The discriminant weights, then, are the elements of the so determined (sometimes normalized) eigenvectors. The variable selection criterion most often employed--implicitly or explicitly--has been the relative size of the so-called "beta weights" [see, e.g., Clemens, et al. (1970)]. However, many other schemes are at our disposal: for example, multiple univariate, stepwise, factor analytic-discriminatory, and correlational procedures. The bases on which some selection criteria are founded are briefly discussed in the next paragraphs.

1) Beta weights. The beta weights are merely the scaled discriminant weights--the multiplication factors being the error standard deviations of the respective variables. A subset of variables is selected by including those variables having large beta weights on the discriminant function(s)--an arbitrary lower bound for the absolute value may be set where a natural cut-off occurs.

2) F-ratios. It has been recommended (e.g., Grizzle, 1970) that to determine which variables ought to be subjected to analysis in the design of an experiment, single variable analyses should be carried out beforehand. The usual univariate analysis performed is a simple ANOVA with the accompanying omnibus F-test; however, if the criterion variable responses are categorical and discrete in nature, the chi-square statistic may be more appropriate. To determine a subset of variables to be used, then, one merely deletes those variables that do not have a reasonable expectation of yielding information about differences

among groups That is, select only those variables that produce significant univariate statistics at, say, the .01 level Or, selection may be based on the relative magnitudes of the statistics themselves.

3) Stepwise values The stepwise discriminant procedure is that outlined in the BMD Manual (Dixon, 1967). At each step that variable is selected i) with the largest value of an F-statistic, or ii) which when partialed on the previously entered variables has the highest multiple correlation with the groups, or iii) which gives the greatest decrease in the ratio of within to total generalized variances Also, a variable is deleted at any step if the value of its associated F-statistic becomes too low. An ordering of the variables is thus determined, and variables may be retained to the point where the value of Wilks' lambda "levels off," or when the increase in the proportion of correct classifications is no longer "appreciable."

4) Component loadings. One factor analytic procedure which may be employed in variable selection is that suggested by Horst (1965, p. 555) It involves a principal component analysis of the matrix of variable intercorrelations (in the present situation an "error" correlation matrix). From a transformation (e.g., varimax) of the pattern matrix, a subset of variables is selected such that each of the components will be adequately represented in the subset Variables are selected which have the highest loadings (in absolute value) on each of the components. Presumably, no variables are selected which have high loadings on more than one component.

5) Factor analytic-discriminatory correlations Bargmann (1962) recommends a maximum likelihood factor analysis (MLFA), with (oblique) rotation to simple structure to define clusters of variables [i.e., factors] which have some

underlying characteristic in common. Once a classification of variables into clusters has been determined, the problem of "representative selection" of variables for the purpose of discrimination may be solved by applying the general discriminatory techniques to each cluster. These techniques involve, for each cluster, finding maximum likelihood estimates of the correlations between the response variables and the artificial variable (the leading "discriminant function") determined by the usual eigenanalysis. Variables are then selected that load on each factor and correlate highly with the respective leading discriminant function. It must be noted that consideration is restricted to only the leading discriminant function since this function usually accounts for a major portion of the discriminatory power of the set of predictors [Bargmann (1970, p. 55) discusses further reasons for considering only the first function in the interpretation of a discriminant analysis]

6) Variable-DF correlations. Some investigators merely order the response variables with respect to contribution to the overall discrimination between the criterion groups by examining estimates of the correlations of the response variable versus the (leading) discriminant function. Selection is then based on these so-called "structure" values without employing any factor analytic techniques. Two approaches have been used to compute these estimates. When the concept of "total population" is meaningful, then the p r-values are determined by the relationship

[1] $$\underline{r} = \underline{v}\, D_{s_1}\, R\, D_{1/s}$$

where

$\underline{v}$ = (1xp) vector of weights for the first discriminant function,

D_{s_1} = (pxp) diagonal matrix of "total" standard deviations of the p variables,

R = (pxp) "total" correlation matrix of the p predictor variables, and

$D_{1/s}$ = (pxp) diagonal matrix of the reciprocal of the standard deviation of the scores on the first discriminant function.

Correlations computed this way are precisely the r-values that would result if the Pearson coefficients were calculated between the sample variable scores and the sample discriminant scores [Gulliksen (1950, p. 339)]. If the concept "k populations equally dispersed" makes sense, then the maximum likelihood estimate of the true correlation vector is given by [Bargmann (1970, p. 53) or Porebski (1966, p. 266)]

[2] $$\underline{r}^* = (\underline{v}\, E\, \underline{v}')^{-\frac{1}{2}} (\underline{v}\, E)\, D_{1/\sqrt{e_{ii}}},$$

where $\underline{v}$ is defined as above,

E = (pxp) pooled within-groups deviation score SSCP matrix, and

$D_{1/\sqrt{e_{ii}}}$ = (pxp) diagonal matrix of the reciprocals of the positive square roots of the diagonal elements of E

As with beta weights, the number of variables in the subset selected may be determined by a drop in the absolute value of the correlation coefficients.

The major purpose in using the selection schemes discussed above is that of determining indicators of variable potency in terms of accuracy of classification. Discriminant analysis employed in this sense is strictly descriptive, or exploratory in that it may provide leads for subsequent investigation

Data

To illustrate the application of the above mentioned selection procedures two sets of data were used. One set involved measures on 13 variables for three samples of college fresmen in a midwestern university for the purpose of assigning them to beginning French courses. Table 1 gives the means and standard deviations on each variable for each of the three groups. A total of 153 (35 + 81 + 37)

-Insert Table 1 About Here-

subjects was involved. Group 1 consisted of those freshmen "correctly" assigned to the beginning French course. Groups 2 and 3 correspond to students in more advanced French courses. The appropriateness of the level of the course in French for assignment of each of the students was determined by teacher judgment after the 1965 fall session ended. More detailed information regarding initial and final group assignment procedures is given by Bisbey (1969).

Seventeen measures on each subject in five educational progress groups of high school students comprise the second set of data. Five small samples randomly selected from a nationwide stratified sample of nearly 26,000 eleventh grade students made up the five criterion groups. The sampling produced a total of 600 (177 + 26 + 75 + 52 + 270) subjects. The five selected samples were also stratified with respect to sex and educational progress. Measures on the 17 predictors were obtained in 1960, while group membership was determined in 1962. Group 1 was comprised of those students who, for two years following high school, had not attended any college. Those students who enrolled in a business college made up Group 2. Group 3 consisted of vocational college students. The other two groups were college enrollees: Group 4 corresponds to students in junior college and Group 5 to senior college students. Means and standard deviations

are presented in Table 2. Variables 1-7 are in the cognitive domain and are measured by various tests of substantial reliability. Measures on Variables 5, 6, and 7 are composites of five, two and three tests, respectively. Variables 8-12

-Insert Table 2 About Here-

are interest variables, and Variables 13-15 are temperament variables. Variable 16, "Curriculum," is a scholasticism variable, measures on which are ordinal with 1=agriculture, . . ., 5=college preparatory. A composite socioeconomic status measure was used for Variable 17.

Results

The results of this study are presented separately for the two sets of data Because the analytic techniques used are exploratory in nature, and for reasons discussed by Nunnally (1967, p. 388) and Porebski (1966, pp. 228-229), the multivariate tests of mean differences and of homogeneity of covariance matrices were not deemed appropriate here.

Three-group case. Values of interest for four of the selection criteria are presented in Table 3. Table 4 gives the orderings of the variables according to four of the criteria employed. Only one "Variable vs DF Correlation" ordering is given since the total group (see equation [1]) and within-groups (see equation [2]) correlations produced identical orderings. Since no values are directly associated with each variable in the stepwise procedure, the resulting ordering is only given in Table 4.

-Insert Tables 3 and 4 About Here-

The correlation matrix that was analyzed by the maximum likelihood and principal axis procedures was the (13 x 13) "error" correlation matrix, R.

The elements of this matrix--"intrinsic" correlations--are given by

$$r_{ij} = \frac{e_{ij}}{\sqrt{e_{ii}}\sqrt{e_{jj}}}$$

where e_{ij} is the (i,j)th element of the matrix E. By using R, spurious correlations or "correlations in widespread classes" are ignored [Bargmann (1968, p. 571)]. To justify the analysis of R, Bartlett's sphericity test using a chi-square statistic [Morrison (1967, p. 113)], which is a measure of the degree to which R differs from I, was performed. The value of the statistic was 742.349, which for df=78, implies a probability, under a true null hypothesis, of less than .001 that this value could be attributed to chance. The factor analysis program[3] employed in this study to obtain, by an iterative procedure, a maximum-likelihood estimate of the factor matrix starts with an improved centroid solution. The built-in test (a chi-square statistic) for the number of significant factors indicated that five factors were adequate. The five resulting factors were then rotated obliquely via MAXPLANE [Eber (1966)] in an attempt to arrive at a simple structure solution. The rotated factor loadings are given on the left in Table 5. The

-Insert Table 5 About Here-

following clusters of variables were determined:

Factor I - Variables 1, 2, 3, 8

Factor II - Variables 5, 8, 9, 10, 11, 12

Factor III- Variables 6, 11, 12, 13

Factor IV - Variables 2, 3, 4, 7, 11

Factor V - Variables 2, 3 4, 5, 6, 7, 8

Five discriminant analyses were then performed using only the variables determining the above factors. Estimates of the correlations between the discriminant functions and the variables involved in the functions were calculated according to equation [2]. The vectors of correlations evaluated for each of the five factors are:

Factor I

$$\underline{r}^* = [.87, .43, .91, .22] \quad (1, 2, 3, 8)$$

Factor II

$$\underline{r}^* = [.48, .03, .15, .15, .67, .84] \quad (5, 8, 9, 10, 11, 12)$$

Factor III

$$\underline{r}^* = [.18, .66, .83, -.22] \quad (6, 11, 12, 13)$$

Factor IV

$$\underline{r}^* = [.07, .20, .16, .31, .84] \quad (2, 3, 4, 7, 11)$$

Factor V

$$\underline{r}^* = [.11, .26, .21, .66, .28, .33, .06] \quad (2, 3, 4, 5, 6, 7, 8)$$

The orderings of the variables determined by the descending order of the absolute values of these correlations are as follows:

Factor I - (3, 1, 2, 8)

Factor II - (12, 11, 5, 9, 10)

Factor III- (12, 11, 13, 6)

Factor IV - (11, 7, 3, 4, 2)

Factor V - (5, 7, 6, 3, 4, 2, 8)

To obtain a representative subset of variables, then, we may choose variables 1, 3, 5, 11, and 12. The number of variables selected by the other five methods was arbitrarily made equal to the number selected here, namely, five, so that a

comparison could be made among the six methods when subsets of the same size are selected.

Two criteria were considered in determining the number of components to extract for the principal axis solution: i) the eigenvalue criterion, and ii) the "scree" test [Cattell (1966, p. 206)]. It was decided to extract four components. The four eigenvalues found were 4.20, 1.97, 1.58, and 1.06 (the fifth was 0.76); 68 percent of total variance was accounted for by the four components. These components were then rotated to meet the varimax criterion. The resulting loadings are given on the right in Table 5. Based on Horst's suggestions, and restricting the number to five, the following variables were selected: 3, 5, 10, 12, and 13.

A summary of the five variables selected according to each of the six criteria presented in Table 6.

The classification scheme that was employed in this study is based on the posterior probability of group membership [see Rule 3 in Huberty and Blommers [preparation (a)]; or Cooley and Lohnes (1962, p. 138)]. The proportions of correct classifications when only those variables selected according to the six criteria are included are given in Table 6. It was of interest to test the hypothesis,

-Insert Table 6 About Here-

that, for a given sample of subjects, the classification accuracy is identical when using only those variables selected according to the various schemes. Inclusion of all 13 variables--which gave a proportion of .941--was done to determine if the discriminatory power is significantly decreased when using only those variables selected. A chi-square statistic (Q) suggested by Cochran (1950) was employed. For the dichotomous data here the χ^2 approximation was judged satisfactory [Tate and Brown (1970)]. The overall test of the differences in frequency of correct

classifications resulted in a chi-square value of 26.588 (6 df), with $p<.0005$. The test of "All 13" versus the six selected subsets yielded a $\chi^2(1) = 12.650$, with $p<.0005$. However, when tests were carried out comparing the accuracy of the selection criteria in a pairwise fashion, not one test statistic was significant. And only the test of "All 13" versus the principal axis results, which yielded a Q-value of 15.385 with a nominal probability of less than .0005, was judged significant.

The number of variables in each subset was determined by the number selected by the method involving MLFA. It may be of interest to investigate the subsets yielded by each criteria without this restriction. If the number of variables to be selected was not limited to five, the following subsets would probably have been chosen using the respective criteria: beta weights -- 2, 3, 4, 5, 7, 9, 10, 11, 12, 13; F-ratios -- 1, 3, 4, 5, 6, 7, 11, 12, 13; stepwise values -- 2, 3, 4, 5, 6, 7, 9, 11, 12, 13; principal axis loadings -- 1, 3, 4, 5, 7, 8, 9, 10, 12, 13; variable-DF correlations -- 1, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13. The corresponding proportions of correct classifications obtained using these subsets were .908, .915, .935, .895, and .928 while the proportion yielded by the subset determined by the method involving MLFA was .882. The classification accuracy yielded by these six subsets along with that yielded by all 13 variables, viz., .941, were not found to be significantly different ($Q = 10.030$, $p > .10$).

Five group case. The same procedures were followed here as in the three group case. Some indices of variable potency for the data involving five criterion groups are given in Table 7. Orderings of the 17 variables based on four criteria are presented in Table 8. Again the two sets of correlations

-Insert Tables 7 and 8 About Here-

of variable scores versus discriminant scores were identically ordered.

Bartlett's test for sphericity yielded a χ^2 (136) = 3030.52, with $p < .001$. Thus it was justifiable to analyze the (17 x 17) error correlation matrix, R, by both the maximum likelihood and principal axis methods. The eight resulting maximum likelihood factors were rotated so that the MAXPLANE criterion was met; the loadings are given on the left in Table 9. Separate discriminant analyses were performed using only those variables determining the eight factors. The "within-groups" correlations of those variables involved against

-Insert Table 9 About Here-

the respective leading discriminant functions were then found. The following correlation vectors resulted:

Factor I

	2	1	7
$\underline{r}^*$ =	[.94	.87	.14]

Factor II

	8	12	10	11
$\underline{r}^*$ =	[.61	-.34	.32	.15]

Factor III

	6	15	13
$\underline{r}^*$ =	[.97	.43	.15]

Factor IV

	6	3	5	7
$\underline{r}^*$ =	[.99	.43	.37	.13]

Factor V

	16	17
$\underline{r}^*$ =	[.99	.51]

Factor VI

	9	12	10	11
$\underline{r}^*$ =	[.59	-.46	.40	.19]

Factor VII

	6	13	14
$\underline{r}^*$ =	[.99	.16	.08]

Factor VIII

	6	8	4	5	12
$\underline{r}^*$ =	[.88	.48	.38	.32	-.25]

Based on these correlations it was decided that variables 1, 2, 6, 8, 9, 12, and 16 could be considered a representative subset. The number of variables chosen on the basis of each of the other selection criteria was subsequently restricted to seven.

The (varimax) rotated principal axis loadings are presented on the right in Table 9. Of the six eigenvalues which accounted for 68 percent of the total variance, five were greater than unity. On the basis of (the absolute values of) the loadings it was decided to chose the following seven variables: 1, 4, 7, 10, 13, 16, and 17.

The seven variables selected, based on each of the six criteria, are indicated in Table 10. As in the three-group case the comparative efficiency of the selection

-Insert Table 10 About Here-

rules was determined by the accuracy in classifying the 600 subjects into their respective groups. The proportion of correct classifications when all 17 variables were considered was .750; proportions for the six subsets determined by the various criteria are given in Table 10.

A highly significant ($p<.0005$) Q-value of 99.642 (6 df) resulted from the overall test of the classification accuracy of the six selected subsets and all 17 variables. The test of "All 17" versus the six subsets yielded a $\chi^2(1) = 6120.214$, with $p<.0005$. It was found that the accuracy yielded by all 17

variables was significantly higher than the accuracy yielded by each of the six selected subsets, while their was no evidence to conclude that the accuracies yielded by the six subsets were different.

Even when the size of each subset was not restricted to seven and selection of subsets was made independently, a significant loss of discriminatory power resulted for each subset. The variables selected by five criteria when the number was not restricted are as follows: beta weights -- 2, 3, 6, 8, 9, 10, 11, 12, 15, 16; F-ratios -- 1, 2, 3, 4, 5, 6, 8, 9, 15, 16; stepwise values -- 2, 3, 4, 6, 8, 9, 10, 11, 12, 16; principal axis loadings -- 1, 2, 4, 5, 7, 10, 12, 13, 15, 16, 17; variable-DF correlations -- 1, 2, 3, 4, 5, 6, 8, 9, 15, 16. The corresponding proportions of correct classifications using these subsets were .637, .657, .642, and .623, while the proportion yielded by the subset determined by the method involving MLFA was .602. Again it could not be concluded that these six subsets differed in classification accuracy.

Discussion

When selection was based on high principal axis loadings the resulting loss in discrimination power was highly significant with both sets of data. That this scheme selected a subset of variables with lower discriminatory power may be expected. As Bargmann (1968, p. 574) states, ". . . it can be argued that variables which have the highest correlation against the first principal component contribute most strongly to discrimination among individual experimental units." In the analysis considered here, however, it is the between-groups variation which is of interest, not the between-units variation.

The popularity of the use of weights applicable to variables in standard form (herein called "beta weights") is probably due to the familiarity of multiple regression techniques to many researchers. The problem of instability of beta weights over repeated sampling exists in discriminant as well as in regression analysis (Huberty and Blommers[in preparation (b)]. Bock and Haggard (1968, p. 118) elaborate on the statement that, "Even when standardized, the coefficients of the discriminant function(s) do not always reflect closely the direction or magnitude of effects in corresponding variables." Further, the relative ordering of variable contribution is not preserved after some variables are deleted; it is necessary to recompute the weights with the fewer number of variables.

On the deletion of nonsignificant variables, Grizzle (1970, p. 319) comments that ". . .the chance of finding significant differences by a multivariate test is lessened when dependent variables which do not contain information about differences among treatments are included in the analysis." It has been argued, however, that even though a variable may produce a nonsignificant univariate statistic, because of the discriminator intercorrelations this variable may appear relevant to the discrimination when included in the whole set of discriminators.

As in regression analysis there are some concerns over the use of the stepwise procedure in discriminant analysis. Two such concerns are i) the estimated ordering of the variables by this method may be biased (because of the presence of "error suppressor" variables), and ii) this method will not necessarily lead to selection of the "best" subset of a given size.

If, in reducing the number of predictors, interest is centered on obtaining a subset that best represents the set of original variables, then the factor analytic-discriminatory method may be the most appealing. Usually, however, the primary concern is one of arriving at a subset which will have nearly the same discriminatory power as the original set; representativeness and high discriminatory power will not necessarily simultaneously characterize the subset chosen.

There are two characteristics of the variable-discriminant function correlations which ought to be mentioned. The first is that the rank-order of these correlations is unchanged after some variables have been deleted. Secondly, the rank-order remains invariant under a monotonic transformation of the variables.

Proportions of correct classifications were obtained for subsets of various sizes determined by four selection criteria to further assess these criteria in terms of the relative discriminatory power of the chosen subsets. Subsets of size three to size p-1 (12 when k=3, and 16 when k=5) were considered. The proportions obtained for the four selection criteria are given in Tables 11 and 12. In the three group case all subsets, with a single exception, determined

-Insert Tables 11 and 12 About Here-

by the stepwise procedure exhibited at least as much discriminatory power as any of the subsets determined by any of the other three criteria. Of the fourteen subset sizes considered in the five group case the stepwise procedure determined the best subset eight times.

When investigating the relative discriminatory power of selected subsets either between or within selection procedures we concede that these procedures do not ordinarily yield the best subset of retained variables of a given size.

Further, retention or deletion of some variables is often made after such considerations as availability, reliability, validity, and cost of measures as well as contribution to discrimination.

Conclusion

Of course, a definitive answer to the question of which of the selection methods studied is "best" cannot be given from the results of such an empirical investigation. Nonetheless, the present study does shed some light on the relative merits of some frequently used methods of determining variable contribution to discrimination. If a single selection method from those studied were to be chosen as best, based on the data used, it would be the stepwise method. There quite possibly is a method that may be superior to any studied here, however. To date, a method of variable selection in regression analysis which appears to have considerable promise is that given by Hocking and Leslie (1967). In searching for the best subset of size q ($<p$) a considerable number of the $\binom{p}{q}$ possible subsets is eliminated. The criterion used in finding the desired subset from those remaining is the reduction in the regression sum of squares due to removing a set of p-q variables. Applying their techniques to discriminant analysis might involve a quadratic form denoted by Rao (1952, p. 257) as "V" and which may be termed a "generalized Mahalanobis D^2." This approach deserves further study. Another possibility which merits consideration is that of determining criteria to be employed in a "backward" selection scheme (in which variables are successively discarded one at a time from the original full set)--this is in contrast to the BMD "forward" procedure.

Since an optimum analytical solution to the variable selection problem in discriminant analysis is not expected to appear on the scene in the very near future, arriving at a selection procedure better than those now available may be possible through a Monte Carlo study. This approach might also be used to arrive at standard errors of the discriminant coefficients or of the variable-DF correlations.

FOOTNOTES

[1]Professor Warren G. Findley is acknowledged for reading an earlier draft of the manuscript.

[2]Some researchers consider discriminatory analysis and (multivariate) classificatory analysis as separate or even independent techniques.

[3]A modified version of FCAN, by Professor R. E. Bargmann, was used.

REFERENCES

Bargmann, R. E. Representative ordering and selection of variables. Final Report, Cooperative Research Project No. 1132, USOE, 1962.

Bargmann, R. E. Exploratory technique involving artificial variables. In P. R. Krishnaiah (Ed.), Multivariate analysis - II (Proceedings of second international symposium on multivariate analysis). New York: Academic Press, 1968.

Bargmann, R. E. Interpretation and use of a generalized discriminant function. In R. C. Bose, et al. (Eds.), Essays in probability and statistics. Chapel Hill: University of North Carolina Press, 1970.

Bisbey, G. D. Use of multiple discriminant analysis techniques in the placement of students in college French. Unpublished doctoral dissertation, University of Iowa, 1969.

Bock, R. D. and Haggard, E. A. The use of multivariate analysis of variance in behavioral research. In D. Whitla (Ed.), Handbook of measurement and assessment in behavioral sciences. Reading, Mass.: Addison-Wesley, 1968.

Cattell, R. B. The meaning and strategic use of factor analysis. In R. B. Cattell (Ed.), Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966.

Clemens, B., et al. Engineers' interest patterns: then and now. Educational and Psychological Measurement, 1970, 30, 675-686.

Cochran, W. G. The comparison of percentages in matched samples. Biometrics, 1950, 37, 256-266.

Cochran, W. G. On the performance of the linear discriminant function. Technometrics, 1964, 6, 179-190.

Cochran, W. G. and Bliss, C. I. Discriminant functions with covariance. Annals of Mathematical Statistics, 1948, 19, 151-176.

Collier, R. O. A note on the multiple regression techniques for deleting variables in the discriminant function. Journal of Experimental Education, 1963, 31, 351-353.

Cooley, W. W. and Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

DeMann, M. A predictive study of rehabilitation counseling outcomes. Journal of Counseling Psychology, 1963, 10, 340-343.

Dixon, W. J. (Ed.) Biomedical computer programs. Los Angeles: University of California Press, 1967. Pp. 214a - 214t.

Dunteman, G. H. Discriminant analysis of the SVIB for female students in five college curricula. Journal of Applied Psychology, 1966, 50, 509-515.

Eber, H. W. Toward oblique simple structure: MAXPLANE. Multivariate Behavioral Research, 1966, 1, 112-125.

Grimsley, G. and Summers, G. W. Selection techniques for Pakistani post graduate students of business. Educational and Psychological Measurement, 1965, 25, 1133-1142.

Grizzle, J. E. Response curves and multivariate comparisons. In R. C. Bose, et al. (Eds.), Essays in probability and statistics. Chapel Hill: University of North Carolina Press, 1970.

Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.

Hocking, R. R. and Leslie, R. N. Selection of the best subset in regression analysis. Technometrics, 1967, 9, 531-540.

Horst, P. Factor analysis of data matrices. New York: Holt, Rinehart and Winston, 1965.

Horst, P. and Smith, S. The discrimination of two racial samples. Psychometrika, 1950, 15, 271-289.

Huberty, C. J and Blommers, P. J. An empirical comparison of selected classification rules, in preparation (a).

Huberty, C. J and Blommers, P. J. An empirical comparison of some indices of variable contribution in multiple group discriminant analysis, in preparation (b).

Kendall, M. G. A course in multivariate analysis. New York: Hafner, 1957.

Morrison, D. F. Multivariate statistical methods. New York: McGraw-Hill, 1967.

Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, 1967.

Quenoville, M. H. Note on the elimination of insignificant variates in discriminatory analysis. Annals of Eugenics, 1949a, 14, 305-308.

Quenoville, M. H. A further note on discriminatory analysis. Annals of Eugenics, 1949b, 15, 11-14.

Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.

Rao, C.R. Linear statistical inference and its applications. New York: Wiley, 1965.

Tate, M. W. and Brown, S. M. Note on the Cochran Q test. Journal of the American Statistical Association, 1970, 65, 155-160.

Vroegh, K., et al. Discriminant analyses of preschool masculinity and femininity. Multivariate Behavioral Research, 1967, 2, 299-313.

Wallace, W. and Travers, R. M. W. A psychometric sociological study of a group of speciality salesmen. Annals of Eugenics, 1938, 8, 266-302.

Table 1

Means and Standard Deviations

Three-Group Case[a]

Variable No.	Name	Means Group 1 (N=35)	Group 2 (N=81)	Group 3 (N=37)
1	High School English Cumulative Grade Point Average	2.91 (0.50)	3.29 (0.50)	3.33 (0.49)
2	High School Mathematics Cumulative Grade Point Average	2.52 (0.70)	2.86 (0.80)	2.69 (0.80)
3	High School Social Science Cumulative Grade Point Average	2.89 (0.48)	3.34 (0.52)	3.21 (0.55)
4	High School Natural Science Cumulative Grade Point Average	2.65 (0.54)	3.07 (0.60)	2.95 (0.69)
5	The Number of Semesters of High School French Taken	3.17 (0.98)	4.23 (0.76)	4.97 (1.12)
6	High School French Cumulative Grade Point Average	2.68 (0.69)	3.20 (0.64)	3.15 (0.67)
7	American College Testing Program--English (Standard Score)	22.43 (2.09)	23.80 (2.86)	25.03 (2.51)
8	American College Testing Program Mathematics (Standard Score)	23.66 (5.73)	24.79 (4.99)	24.27 (5.35)
9	American College Testing Program Social Sciences (Standard Score)	25.14 (3.96)	26.59 (3.14)	27.32 (3.68)
10	American College Testing Program Natural Sciences (Standard Score)	23.31 (4.16)	24.63 (4.57)	26.05 (4.61)
11	The ETS Cooperative French Placement Test Score--Aural Comprehension	14.34 (4.78)	19.27 (3.62)	25.27 (3.96)
12	The ETS Cooperative French Placement Test Score--Grammar	40.86 (7.55)	52.88 (7.22)	63.76 (4.01)
13	The Number of Semesters Since Last High School French Course Was Taken	2.00 (2.16)	1.40 (1.68)	0.49 (0.99)

[a]Standard deviations in parentheses

Table 2

Means and Standard Deviations

No.	Variable Name	Five-Group Case[a] Means Group 1 (N=177)	Group 2 (N=26)	Group 3 (N=75)	Group 4 (N=52)	Group 5 (N=270)
1	Literature Information	11.75 (4.08)	13.12 (2.94)	13.53 (4.59)	14.38 (4.02)	15.94 (4.53)
2	Social Studies Information	13.71 (5.16)	14.50 (4.64)	15.37 (5.11)	17.44 (4.70)	18.56 (4.45)
3	English Total	79.56 (19.38)	83.69 (19.00)	81.39 (14.97)	85.56 (15.49)	89.36 (15.23)
4	Mechanical Reasoning	9.58 (4.13)	8.81 (4.35)	12.37 (4.03)	10.83 (4.31)	12.35 (4.43)
5	Visualization in 3 Dimension	8.05 (3.32)	7.42 (3.46)	9.56 (3.19)	8.29 (3.33)	7.79 (3.40)
6	Mathematics Information	11.53 (5.55)	11.23 (3.85)	13.63 (7.22)	15.60 (6.88)	20.39 (7.73)
7	Clerical-Perceptual Speed	74.14 (28.09)	71.58 (28.71)	76.71 (25.67)	71.92 (23.36)	78.31 (24.93)
8	Physical Science Interest	13.31 (8.32)	10.73 (7.64)	15.53 (8.64)	15.98 (7.92)	19.67 (10.07)
9	Literary-Linguistic Interest	16.81 (8.66)	17.62 (8.85)	15.43 (10.52)	17.29 (9.00)	21.10 (9.30)
10	Business Management Interest	16.23 (7.62)	15.19 (5.50)	15.17 (8.47)	15.65 (7.29)	18.58 (8.21)
11	Computation Interest	14.92 (8.89)	13.50 (7.97)	13.36 (8.37)	14.15 (8.18)	15.83 (9.27)
12	Skilled Trades Interest	11.03 (6.58)	7.50 (3.98)	11.63 (7.66)	9.27 (6.42)	8.64 (6.15)
13	Sociability	6.33 (3.05)	6.42 (2.67)	6.31 (3.03)	6.40 (3.17)	6.93 (2.94)
14	Inpulsiveness	1.94 (1.44)	1.85 (1.87)	1.87 (1.69)	1.70 (1.44)	2.07 (1.80)
15	Mature Personality	10.88 (4.78)	10.35 (3.64)	11.35 (5.33)	12.35 (4.83)	13.76 (5.57)
16	Curriculum	3.33 (1.42)	3.77 (1.11)	3.42 (1.61)	3.98 (1.50)	4.30 (1.55)
17	Socioeconomic Composite	92.82 (18.13)	96.81 (21.18)	94.96 (17.53)	98.54 (15.24)	100.20 (24.31)

[a]Standard deviations in parentheses

Table 3

Partial Data for the Three-Group Case

Variable	Beta Weight	Univariate F-ratio	Variable vs DF Correlation	
			Total Group	Within Groups
1	0.43	8.50**	0.34	0.18
2	-2.46	2.40	0.10	0.05
3	2.02	9.28**	0.26	0.14
4	6.18	5.84**	0.21	0.11
5	3.42	35.60**	0.66	0.41
6	0.30	8.23**	0.29	0.16
7	2.18	8.84**	0.37	0.20
8	0.84	0.58	0.05	0.03
9	-2.62	3.72*	0.25	0.13
10	-1.74	3.36*	0.24	0.12
11	6.43	67.87**	0.78	0.56
12	7.28	105.86**	0.88	0.70
13	1.87	7.97**	-0.35	-0.19

* $p < .05$

** $p < .01$

Table 4

Ordering of Variables

Three-Group Case

Beta Weights	Univariate F-ratios	Stepwise Values	Variable vs DF Correlations
12	12	12	12
11	11	11	11
4	5	4	5
5	3	5	7
9	7	3	13
2	1	9	1
7	6	7	6
3	13	6	3
13	4	13	9
10	9	2	10
8	10	10	4
1	2	8	2
6	8	1	8

Table 5

Rotated Loadings for the Three-Group Case[a]

	Maximum Likelihood					Principal Axis			
Variable	I	II	III	IV	V	I	II	III	IV
1	80	07	03	-13	-01	76	11	16	26
2	-42	12	01	-44	-64	66	40	-19	10
3	20	11	05	-80	25	85	02	-02	07
4	-17	13	-02	-75	-26	80	24	-18	14
5	00	33	06	-15	69	-24	11	23	-80
6	-01	-02	31	-05	-65	66	19	21	43
7	08	08	12	66	-90	07	52	14	64
8	-46	31	-10	01	-72	21	70	-28	18
9	03	61	09	-09	-02	24	73	16	-09
10	-03	74	-02	07	-14	08	85	01	05
11	-01	34	31	45	07	-44	34	49	-07
12	01	25	46	13	18	-02	12	74	-21
13	05	07	-59	-06	-05	-08	30	-75	-18

[a]Decimal points omitted

Table 6

The Five Variables Selected

and Proportions of Correct Classifications

Three-Group Case

Beta Weights	Univariate F-ratios	Stepwise Values	Principal Axis Loadings	ML Loadings vs DF Correlations	Variable vs DF Correlations
4	3	3	3	1	5
5	5	4	5	3	7
9	7	5	10	5	11
11	11	11	12	11	12
12	12	12	13	12	13
Proportion:					
.876	.899	.895	.810	.882	.869

Table 7

Partial Data for the Five-Group Case

Variable	Beta Weight	Univariate F-ratio	Variable vs DF Correlation	
			Total Group	Within Groups
1	-6.62	26.15**	0.61	0.53
2	36.60	30.25**	0.66	0.57
3	-24.78	10.20**	0.41	0.33
4	-2.09	14.60**	0.38	0.32
5	-3.69	9.90**	0.33	0.27
6	91.02	51.52**	0.82	0.76
7	0.09	1.28	0.11	
8	36.12	16.38**	0.49	
9	37.13	9.13**	0.37	
10	44.93	4.84**	0.26	
11	-30.83	1.53	0.11	
12	-65.30	6.36**	-0.27	
13	3.27	1.41	0.15	
14	-4.81	0.73	0.07	
15	9.15	10.21**	0.41	
16	19.19	13.41**	0.46	
17	-1.75	3.55**	0.24	

**$p < .01$

Table 8

Ordering of Variables According

to Four Criteria

Five-Group Case

Beta Weights	Univariate F-ratios	Stepwise Values	Variable vs DF Correlations
6	6	6	6
12	2	2	2
10	1	4	1
9	8	9	8
2	4	12	16
8	16	10	15
11	15	3	3
3	3	8	4
16	5	11	9
15	9	16	5
1	12	1	12
14	10	5	10
5	17	7	17
13	11	15	13
4	13	14	11
17	7	17	7
7	14	13	14

Table 9

Rotated Loadings for the Five-Group Case[a]

	Maximum Likelihood								Principal Axis					
Variable	I	II	III	IV	V	VI	VII	VIII	I	II	III	IV	V	VI
1	67	-12	11	-11	07	-05	05	-01	16	-12	09	10	09	87
2	73	-01	11	00	01	04	04	10	33	-05	11	04	-09	77
3	06	-06	19	-50	06	-02	-08	06	43	-12	11	06	57	37
4	-05	05	14	-10	00	12	00	81	85	10	-01	-05	01	15
5	-13	06	06	-29	08	11	03	58	76	05	-02	00	20	07
6	13	09	31	-28	06	-00	-30	32	67	09	08	19	16	37
7	-28	-02	08	-47	12	-04	02	01	22	-06	06	10	81	-14
8	11	47	09	15	-06	-14	-11	45	41	68	-03	00	-15	16
9	06	17	07	-10	04	-43	09	-16	-39	53	05	08	41	36
10	02	66	01	11	-01	-27	08	01	-11	83	11	03	-01	-05
11	-10	58	16	07	-02	-15	-19	02	09	72	11	13	04	-16
12	-08	48	-16	07	-12	-23	-10	33	09	74	-09	-16	-04	-10
13	-14	07	50	05	05	03	34	-02	-04	01	84	00	13	-05
14	-01	00	06	-13	10	-10	40	10	09	08	35	-11	29	22
15	08	10	59	14	02	06	08	-01	12	04	76	11	-07	20
16	-00	00	-01	-02	86	04	-01	-04	00	05	04	87	03	11
17	-10	-09	00	-08	61	-04	07	06	04	-02	00	84	06	02

[a]Decimal points omitted

Table 10

The Seven Variables Selected

and Proportions of Correct Classifications

Five-Group Case

Beta Weights	Univariate F-ratios	Stepwise Values	Principal Axis Loadings	ML Loadings vs DF Correlations	Variable vs DF Correlations
2	1	2	1	1	1
6	2	3	4	2	2
8	4	4	7	6	3
9	6	6	10	8	6
10	8	9	13	9	8
11	15	10	16	12	1
12	16	12	17	16	16
Proportion:					
.618	.622	.625	.577	.602	.612

Table 11

Proportions of Correct Classification

Three-Group Case

Number of Variables	Beta Weights	Univariate F-ratios	Stepwise Values	Variable vs DF Correlations
3	.876	.830	.876	.830
4	.882	.882	.882	.882
5	.867	.889	.895	.869
6	.869	.889	.908	.889
7	.895	.889	.902	.902
8	.908	.889	.908	.889
9	.908	.915	.908	.902
10	.908	.922	.935	.915
11	.928	.928	.928	.928
12	.941	.922	.948	.922
13	.941	.941	.941	.941

Table 12

Proportions of Correct Classifications

Five-Group Case

Number of Variables	Beta Weights	Univariate F-ratios	Stepwise Values	Variable vs DF Correlations
3	.583	.575	.575	.575
4	.588	.600	.585	.600
5	.595	.599	.592	.593
6	.592	.607	.603	.610
7	.618	.622	.625	.612
8	.615	.608	.622	.608
9	.627	.617	.653	.623
10	.637	.623	.657	.623
11	.650	.647	.670	.647
12	.685	.669	.672	.668
13	.697	.682	.677	.682
14	.705	.705	.713	.698
15	.713	.710	.747	.710
16	.728	.733	.742	.733
17	.750	.750	.750	.750